15.458 – Financial Data Science and Computing

Course Description. First in a two-course sequence that covers methods of managing and extracting information from financial data.

To understand modern financial markets and corporate decision making, it is essential to understand financial data: where it comes from, what it conveys (as well as what it hides), and how to analyze it in diverse contexts such as accounting, trading, planning, investing, portfolio management, risk management, corporate finance, efficient markets, and adaptive markets. Financial data sets are often enormous in size and complexity, and many problems require efficiently merging, filtering, and slicing multiple sources at once. In these courses we'll look at this data up close. Students will learn how data relationships are structured and how to use modern tools and technologies to manipulate, clean, manage, and analyze it. Projects will be able to draw from extensive commercial and research data sources via a dedicated high-performance computing platform. At the end of the sequence students will have practical knowledge and experience that will be of value in other courses at Sloan and in industry as well. The course will use real-world data, applications, and cases to illustrate the principles and provide hands-on experience. Finance is an empirical subject, and possessing a practical mastery of financial data will give you an edge.

Pre-requisites and complements. This course is geared toward M.Fin. students. Others may enroll with the pre-requisite of 15.401, or its equivalent, and the educational background of students in the M.Fin. program. As this course provides tools and techniques for real-world applications of finance theory, it is highly complementary to 15.433, 15.437, 15.450, and 15.456; and it is recommended for 15.487, "Algorithmic Trading and Quantitative Investment Strategies" and 15.S08, "Advanced Data Analytics and Machine Learning in Finance." Basic programming skills will be necessary along with familiarity with Excel. Projects will make use of R and SQL, although prior experience in these languages is not required. Students who have not had significant experience with R should plan to complete an online tutorial at the start of the course.

Course Requirements and Grading. Course requirements include: regular attendance and class preparation/participation in lectures and recitations (10 percent) and project/problem sets (90 percent). Data science involves extracting and communicating insights from observational data, and assignments will be evaluated accordingly, with this approximate breakdown: technical quality (50%), analytical quality (25%), and presentation quality (25%). The major assignments are *tentatively* planned as follows:

Project	*Date posted*	*Date due*
A	*September 6*	*September 14*
B	*September 14*	*October 2*
C	*October 2*	*October 18*

Values @ Sloan. Students should be familiar with the detailed guidelines in the Values@Sloan document, attached for reference to the end of this syllabus, which will be strictly followed. Team projects will adhere to the rules of "Type 3" collaboration. Each student must make a substantial contribution to the deliverables and each student will be responsible for the quality and integrity of the project. You are permitted to study with other students currently taking 15.458 to learn about, explore, and master the technology and the various data sources used in the class. However, all work done on assignments must be entirely by your own team. If you have any questions about how this policy applies, be sure to ask.

Course Materials. Lecture notes, handouts, projects, etc. will be posted on Canvas. Data sets for problems, projects, and research will be available on the class database server, obelix.mit.edu. Textbooks and readings assigned for the course will be posted to Canvas, as will pointers to software, tutorials, etc. during the course. Use of any materials from other courses not taken by you is prohibited including, but not limited to, problem sets, solution sets, lecture notes, code, spreadsheets, or data, whether created by an instructor, a student, or anyone else.

Class Preparation and Participation. Class preparation and participation are important components of this course. This course is "hands on," so students should be ready to reproduce the material and cases presented in lecture on their own. Students are expected to come to each class well prepared to discuss their results from previous and current assignments.

Computing Resources. Computing and data resources will be introduced in lecture and in recitation. Students can make use of the virtual lab (which has replaced the former Sloan trading room), the Bloomberg terminals, and resources on the MIT network, along with MIT's access to multiple data vendors. A dedicated set of data servers and a data warehouse are available for class projects, and students are encouraged to familiarize themselves with these tools: Excel, Access, Matlab (all available from IS&T), R and RStudio (open-source, available via links on Stellar), Microsoft's SQL Server for its query and data management tools (Express and Developer editions available free from Microsoft), and ERwin for its data modeling tools.

Course times and locations:

Meeting	*Lecture*	*Recitation*
Time	T/Th 10-11:30am	F 10-11am
Location	E51-345	E51-151

Course staff:

Name	*Role*	*E-mail*	*Phone*	*Office + hours*
Paul Mende	Instructor	mende@mit.edu	(617) 715-4835	E62-612 Th 12:30-2pm
Youssef Berrada	TA	yberrada@mit.edu	(202) 468-2571	TBA
Robert Kissinger	Admin. support	rkissing@mit.edu	(617) 253-7006	E62-671

15.458 / 15.459 List of Topics

The course will draw from the list of topics below. Note that not all will be covered.

1. Mathematical, Financial, and Technical Preliminaries

a. Course overview, and the role of data in quantitative finance
b. Data technologies past, present, and future
c. Data and computational architecture of a quantitative strategy
d. From risk measurement to risk management

2. Sources of Financial Data

a. Markets and exchanges
b. Corporate data
c. Financial analysts
d. Time series data
e. Asset attributes and corporate actions
f. Economic data
g. Data vendors and data delivery
h. Time scales, scaling, and irregularly spaced data
i. Metadata and meta-metadata
j. Slowly changing data
k. Data integrity; error detection and correction; audit trails

3. Mathematics and statistics in financial analysis

a. Brief tour of stochastic processes, financial time series, and portfolio theory
b. Descriptive statistics
c. Exploratory data analysis
d. Hypothesis testing and statistical inference
e. Regression models
f. Time series models applied to financial data
g. Cross-sectional predictions and time-series forecasting

4. Data Architecture and Design

a. Logical and physical architectures
b. Transactions, operations, decision support
c. Entity-relationship modeling
d. Dimensional design
e. Data design for financial analytics
f. Integrating data from multiple sources; security masters
g. Data domains: market, model, portfolio

5. Data Management Tools and Technologies

a. Scale, performance, and dealing with massive data sets
b. Single-user systems: Excel, Access, …
c. Client-server systems: RDBMS
d. Multi-dimensional databases, OLAP, "business intelligence"
e. Big Data tools, technologies, and approaches
f. Data extraction, transformation, and loading
g. Query design and optimization
h. Data visualization tools

6. Communicating with Data

a. Forms of data presentation
b. Designing storage and analytics with presentation in mind
c. Reports, dashboards, and interactive data exploration
d. Graphics and the structure of graphical data design
e. Precision, accuracy, and units

7. Varieties of Time Series Data

a. Regularly and irregularly spaced data
b. Analytics and scaling
c. Prices
d. Returns
e. Trades
f. Quotes
g. Volumes
h. Derived and calculated data
i. Dealing with redundancy and correction
j. Time series of windowed data (e.g., volatility)

8. Applications

a. Portfolio management
b. Algorithmic trading
c. Quantitative equity strategies
d. Option hedging
e. Index arbitrage
f. Convertible arbitrage
g. Risk management and risk aggregation
h. Performance attribution and factor analysis
i. Volatility surfaces
j. Volatility modeling and trading
k. Dispersion

9. Digital Dashboards

a. Data visualization

b. Taxonomy of graphs for financial analysis
c. Static and batch reporting
d. Dynamic graphs and tables
e. Pivot tables and pivot charts
f. Systems for real-time trading

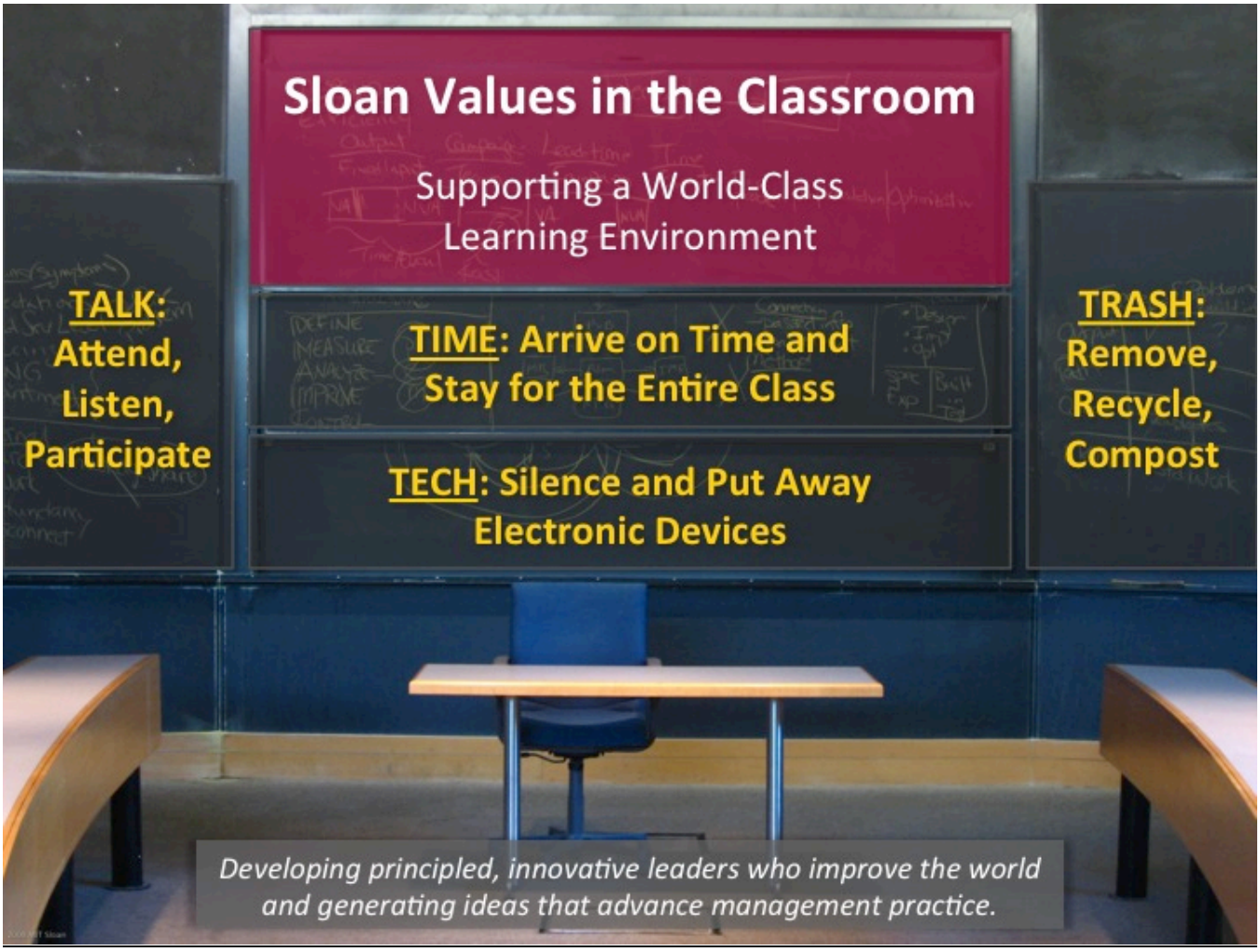
10. Data Aggregation and Data Slicing

a. Factor exposures
b. Risk management
c. Performance attribution
d. Risk factors
e. Industry membership
f. Asset class

11. Machine Learning

a. Supervised and unsupervised learning
b. Use and preparation of data
c. Predictive modeling
d. Reinforcement learning
e. Text mining and natural-language processing

Classroom Values@MITSloan 2015



The image above illustrates some of the Sloan Values that support our learning environment. MIT and Sloan also have a number of policies related to classroom conduct and academic integrity outlined below.

It is your responsibility to make yourself aware of MIT's rules of academic integrity and personal conduct and to adhere to them. When students are found to have violated academic, personal, or professional standards, they are subject to disciplinary action. Possible consequences include grade reduction, an F grade, a transcript notation, delay of graduation, or expulsion from MIT.

MIT SLOAN CLASSROOM CONDUCT POLICIES AND NORMS

MIT Sloan Policy on Classroom Behavior

In order to create a productive learning environment and to ensure mutual respect it is essential that the norms and rules of classroom etiquette and behavior reflect the highest standards. It is also important that these norms be consistently enforced by the faculty across all classes. Although in the final analysis each faculty member is responsible for his or her own classroom, there are significant negative consequences for other faculty and for the School if rules are not consistent and are not enforced. Therefore it is the policy of the MIT Sloan School that

- Students are expected to arrive promptly on time and to stay for the entire class.
- Faculty are expected to begin and end class on time.
- Laptops and e-readers are not be open in the classroom except with explicit permission of the faculty (e.g., when used as part of the instructional program or when required by students because of physical or other challenges)
- Cell phones and PDAs are not to be used or permitted to ring in the classroom.
- Students are expected to attend all classes.

It is expected that faculty will articulate how these rules apply in their class as well as how the rules will be enforced.

A Note re Recruiting

Please note that in accordance with this policy, MIT Sloan requires that students schedule campus interviews outside of scheduled class times and to make every attempt to schedule second round interviews and site visits outside of class times. Classes missed for such activities are not excused absences and may count against your participation grade.

Additional Norms and Expectations

MIT Sloan has developed a number of additional norms that support a healthy learning environment. These include the following:

- Use your plastic nameplate at all times.
- Remain seated for the duration of the class unless it is absolutely essential that you leave, in which case make every effort not to disrupt the class.
- Participate actively and thoughtfully. Quality of contribution matters as much as quantity.
- Allow others to participate – be careful not to dominate conversation.
- Recognize others' contributions in the discussion. Use language such as, "to build on what Pat has just said…."
- When disagreeing, do so respectfully. Use language such as, "I'd like to offer a different perspective…." Acknowledge the possibility that multiple perspectives on complex problems can be valid.
- Be sensitive to the potential impact of comments and behavior on all members of the community. Be open to perspectives of others who come from different personal, cultural, or professional backgrounds from your own.
- Refrain from sidebar conversations.

- If you bring food into the classroom, be sensitive to the noise you are generating. Be sure to remove all trash and dispose of it appropriately (using trash, recycling, and composting facilities).

ACADEMIC HONESTY

As a member of the MIT Sloan academic community, you are expected to uphold the highest standards of academic integrity. Violations of academic integrity include, but are not limited to, cheating, plagiarism, unauthorized collaboration, and facilitating academic dishonesty. These standards are also discussed below, specifically regarding plagiarism, individual work, and team work.

This discussion of academic integrity below is not exhaustive, and there may be areas that remain unclear to you. ***If you are unsure whether some particular course of action is proper, it is your responsibility to consult with your professor and/or teaching assistant for clarification.***

Plagiarism
Plagiarism occurs when you use another's intellectual property (words or ideas) and do not acknowledge that you have done so. Plagiarism is a very serious offense. If it is found that you have plagiarized -- deliberately or inadvertently -- you will face serious consequences, as indicated above.

The best way to avoid plagiarism is to cite your sources - both within the body of your assignment and in a bibliography of sources you used at the end of your document.

Materials gathered through research via the Internet must be cited in the same manner as more traditionally published material. Lack of such citation constitutes plagiarism.

Please see the document Academic Integrity at the Massachusetts Institute of Technology: A Handbook for Students for further discussion of this topic.

Individual Assignments
Many assignments in MIT Sloan courses are expected to be done individually. The information below outlines what is meant by "individual" work. These rules should be observed unless otherwise defined by the instructor.

When you are asked to do ***individual*** work, you are expected to adhere to the following standards:

- Do not copy all or part of another student's work (with or without "permission").
- Do not allow another student to copy your work.
- Do not ask another person to write all or part of an assignment for you.
- Do not work together with another student in order to answer a question, or solve a problem, or write a computer program jointly.
- Do not consult or submit work (in whole or in part) that has been completed by other students in this or previous years for the same or substantially the same assignment.
- Do not use print or internet materials directly related to a case/problem set unless explicitly authorized by the instructor.
- Do not use print or internet materials without explicit quotation and/or citation.
- Do not submit the same, or similar, piece of work for two or more subjects without the explicit approval of the two or more instructors involved.

Please note that many classes will require a combination of team work and individual work. *Be sure that you follow all the guidelines for individual work when a faculty member identifies an assignment as an individual one.*

Team Assignments

When you are asked to ***work in teams***, there is a broad spectrum of faculty expectations. Three general types of appropriate collaboration on team assignments are described below. The instructor will indicate in the syllabus what his/her expectations are. If there is any uncertainty, it is the student's responsibility to clarify with the professor or TA the type of team work that is expected.

Type 1 collaboration:

The professor states that collaboration is allowed, but the final product must be individual. An example of this might be a problem set.

- You are allowed to discuss the assignment with other team members and work through the problems together.
- What you turn in, however, must be your own product, written in your own handwriting, or in a computer file of which you are the sole author.
- Copying another's work or electronic file is not acceptable.

Type 2 collaboration:

The professor states that collaboration is encouraged but that each person's contribution to a given deliverable does not have to be substantial (allowing groups to take a "divide and conquer" approach). An example of this might be a brief progress report that is part of a more extensive collaboration (as a whole, the more extensive collaboration may be Type 3).

- Each team member is encouraged to contribute substantially to the team assignment, however, the team may choose to assign one or more team members to prepare and submit the deliverable on behalf of the team.
- Regardless of how work is shared or responsibilities are divided among individual team members, each member of the team will be held accountable for the academic integrity of the entire assignment. If, for example, one member of the team submits plagiarized work on behalf of the team, the entire team will be subject to sanctions as appropriate.
- The team may not collaborate with other students outside of the team unless the professor explicitly permits such collaboration.

Type 3 collaboration:

The professor states that collaboration is expected and that each team member must contribute substantially to the deliverable. An example of this might be the 15.311 OP project.

- Each team member must make a substantial contribution to the assignment. It is not, for example, acceptable to divide the assignments amongst the team members (e.g., part of the team completes the OP Project while the rest of the team prepares a team case for DMD), though the team may divide the work of any one assignment to complete it as they deem appropriate.
- The team may not collaborate with other students outside of the team unless the professor explicitly permits such collaboration.

If you are unsure whether some particular form of interaction is proper in individual or team work, it is your responsibility to consult the instructor and/or teaching assistant for clarification and guidance.